Jordan Bross. Community, Collaboration and Contribution: Evaluating a BitTorrent Tracker as a Digital Library. A Master's paper for the M.S. in L.S. degree. December, 2010. 40 pages. Advisor: Gary Marchionini.

This study explores the BitTorrent file sharing technology as a means for the construction of a digital library by a community of decentralized internet users. The private BitTorrent tracker What.CD, which currently has over 100,000 users, is the primary focus of the paper. In addition to an overview of the BitTorrent protocol, a detailed analysis of the facets and features of What.CD serves to demonstrate how the collaboration of users contributing their own content via a file sharing website can build an unparalleled library of digital media. It examines the role of authority and privilege at a private BitTorrent tracker, as well as the methods employed by What.CD to maintain a large and well-structured catalog of content produced by a community of anonymous users around the globe.

Headings:

BitTorrent

Digital file sharing

Digital libraries

Peer-to-peer file sharing

COMMUNITY, COLLABORATION AND CONTRIBUTION: EVALUATING A BITTORRENT TRACKER AS A DIGITAL LIBRARY

by
Jordan Bross

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

December, 2010

Approved by:

_____

Gary Marchionini

TABLE OF CONTENTS

Chapter 1: Introduction to What.CD

Over the past two decades, the growth of the Internet as a place where people with shared interests could interact has completely changed the landscape of human collaboration. At the same time, access to information in all forms has become easier than ever. Jeffrey Pomerantz and Gary Marchionini classify digital libraries as an extension of systematic access to information resources, the ideas represented by those resources, and sets of human stakeholders (506). In addition to being a digital extension of traditional libraries, digital libraries have the added value of enabling these stakeholders to reach and influence even broader audiences (Witten et al. 124). From cataloging photographs -- whether digitized or born-digital -- to Google's extensive e-book archives, to museum collections displayed online for the world to visit and browse, digital libraries have brought information to people in new ways.

At the same time, sharing information in the form of data between Internet users has also seen enormous developments during this period. The ability to access files of all types, especially multimedia, made available through the use of various computer programs has become a primary use of the Internet for millions of people. As this form of sharing has evolved, better ways of providing access to, and information about the data being exchanged have arisen. One of the most common methods at present is the use of a file sharing protocol called BitTorrent.

This paper will provide an insight into how a private community, through the use of BitTorrent, has been able to build and maintain one of the most extensive collections of digital music in the world.

With its success in creating this collection, the website What.CD has become an excellent example of how democratic production and management have been the primary tools of a new type of digital library, relying on anonymous participation and collaboration (O'Mahoney 1080). Serving as a digital library and archive consisting of terabytes of digital music files, What.CD is developed and maintained solely by the collaboration and contribution of its users. There have been several studies of the technical aspects of the BitTorrent protocol. This paper avoids an overly mechanical study of this technology, as well as the legal implications of file sharing or copyright. Instead, this paper intends to examine the participatory qualities of a private, "gated" BitTorrent community of users.

The primary use of What.CD for most users is as a website to download music without requiring much more than a decent Internet connection. What.CD does not charge any "real world" currency for access. Not only is the music at What.CD free, but it has a more comprehensive selection of artists and albums, in more file formats and qualities, than any other service, free or paid, anywhere else. Furthermore, it is perhaps better cataloged and organized, and more easily accessible to its users than any other repository of music as well. What.CD is unique in its breadth of material. Ultimately, What.CD demonstrates that a torrent tracker functioning as a digital library becomes a meeting place of users, who are the creators of the library structure, collection developers and curators.

Unlike most repositories, digital or physical, What.CD's primary responsibility is to maintain a well-cataloged and organized system for finding content. Free of the need to pay for massive quantities of hard drive storage and server bandwidth, as well as other features of legal and/or public media websites such as advertising and licensing fees, What.CD can dedicate nearly all of its resources to maintaining and building its collection, as well as expanding and improving features of the site for its users. Compared to early file-sharing services, such as Napster or Kazaa, torrent trackers allow for much more persistent, reliable access to files. Far less dependent on a single user being present and connectable, torrent trackers allow files and data to be better cataloged and retrieved.

Viewing What.CD as a place or destination throughout the course of this paper will help to get a better sense of the way in which the user community behaves and interacts with the site as a digital repository, archive, and library. In their article on digital libraries, authors Pomerantz and Marchionini explore the characteristics of digital and physical libraries. The article makes a variety of observations of the ways in which the traditional, physical library and the digital version differ in facets and uses and also where they (perhaps surprisingly, at times) converge in features and abilities. Essential to both types of collections is that the larger and more vast the quantity of materials, the greater the need for maintenance and organization (Pomerantz and Marchionini 516). This problem is more easily remedied in the virtual space due to the potential for nearly unlimited numbers of users, located in all corners of the world, to work at maintaining a digital library. This feature also solves the dilemma of physical storage space and preservation. By locating files and digital objects in a decentralized network of

individual users' hard drives, storage capacity is significantly increased and potential for the loss of information is significantly decreased (Pomerantz and Marchionini 509).

A brief overview of What.CD – its history, scope and purpose – is first necessary to present the full picture of how the site functions as a digital library. The roots of What.CD can be found in October of 2007. When the largest music torrent tracker at the time, "Oink.me," was forced to close as a result of legal action against it, several new music torrent trackers quickly sprung up in an attempt to fill the void left by Oink.me's absence, and to try and take on the refugees looking for a new music tracker destination. A handful of brand new torrent trackers were created within days of Oink.me closing forever. Small groups of intelligent, resourceful programmers and system engineers worked as quickly as possible to become the "new Oink." After the dust had settled in just a few short weeks, two sites were left that seemed to have the resources and momentum to continue building where Oink had left off.

At present, What.CD has a maximum user capacity of 150,000. There are currently around 125,000 registered users, with approximately 35,000 users active on the site on a daily basis. The amount of individual BitTorrent files available on the site at present is 814,236. This breaks down to 379,267 individual albums, by 310,907 different artists, and this count increases every minute. In comparison, Waffles.fm, the other major music torrent tracker to have originated in the wake of Oink.me's closure (and currently the second largest private music torrent tracker) has 63,000 members, and approximately 300,000 BitTorrent files. There are several explanations for why What.CD has come to be bigger and more comprehensive than Waffles.fm, which speaks to the organization and effort the site continues to possess. While it may have been as simple as

an ability to maintain the servers and keep the site up and reliable in the early days of friendly competition between What.CD and Waffles.fm, it is more likely the variety of features and capabilities of use that What.CD developed over time. Many of these features will be discussed later in this paper. In addition to competing with other torrent trackers, however, it is important not to forget that while it is free, there are many other places on the internet where music can be obtained, both free and paid.

Among several reasons What.CD users prefer the torrent site to other free and paid music content websites is the wide range of options in the music selection. Unlike most paid service, none of the music is encumbered with DRM (Digital Rights Management). Several versions of an album, some released only in select regions of the world, are not available to purchase even if users would like to. Other album variations, in the form of re-releases, deluxe editions, limited editions, promotional copies, etc. not made available through paid services at all in many cases can be found on What.CD. Even more importantly, out-of-print, rare, and unreleased music is readily available on What.CD, in many cases making it the only place in the world to find certain albums or songs.

The benefits of a website such as What.CD go far beyond merely being able to download music for free. All of the music found on What.CD is decentralized, disseminated and held in thousands if not millions of locations around the world, and does not exist physically solely in any one place. As a result, the music can arguably never be lost; it is a digital archive of the world's music collection.

Chapter 2: Overview of BitTorrent Technology

The technology at the center of the What.CD website is BitTorrent. BitTorrent is a file sharing protocol that is used to distribute many types of data:

> The main goal of BitTorrent is to enable scalable content distribution. To this end, the load of distributing a file is shared between the content publisher and those who download it: the peers downloading and those which have already downloaded the file supply bandwidth, the parts of the file they already have, and content availability. (Andrade et al., "Resource Demand" 516)

The use of BitTorrent pertaining to this paper is its use in bundling together sets of mp3 files. A BitTorrent file is extremely small in size, though this differs based on how much actual data the torrent file points to. For example, a collection of 10 mp3 files, each about three megabytes would make a torrent file of under one megabyte. When thinking about the technology behind torrents, it is important to understand that the torrent itself is merely metadata, and is not a container for any other files (Jones, "Strategies and Technologies" 657). A website that hosts these .torrent files, organizes them and makes them available for download is called a tracker. Anna Satsiou and Leandros Tassiulas defines a tracker as:

> …a centralized entity…which keeps track of all the peers who download files specified in the torrent. Each peer that wants to download a file, finds the torrent of interest and connects to the associated tracker, who is responsible to return a random set of peers (called its neighbors) currently transferring pieces of the file(s) specified in the torrent. (468)

In turn, the tracker that catalogs and identifies the torrent files does not store audio files (or any files other than .torrent files) on any server. This is key, because it allows the tracker to store immense amounts of metadata without the need for the vastly greater storage space that would be necessary to house the audio files. Further, it helps to safeguard in the case that the tracker goes down. In this case, it is much more simple to recover the directions which point to the audio files (.torrent metadata) than to recover the files themselves.

Once a user obtains a torrent file (".torrent") they would then use a BitTorrent client (a program/application) -- of which there are many options -- to open the file. The torrent file contains metadata, which points to a "tracker" (the website which hosts information about where pieces of the files, in this case mp3s, are located) and the transfer of small pieces of data that make up those mp3 files begins. The tracker in this case is What.CD. As a result of this setup, no actual music file or copyrighted material is on any torrent tracker. The data that make up the music file or data that the torrents point to are distributed between users. As long as at least one user with the data has his BitTorrent client open, and the specific torrent in said client, the data are available to others. The user(s) who are making the data available via their open client are called "seeders." The users downloading the files from seeders are called "lechers." A successful, completed leech from the a pool of seeders is called a "snatch."

BitTorrent technology and applications are at the core of what makes a site like What.CD function. Without having to host actual music files on its servers, What.CD bypasses copyright law in many countries. While music is being shared between users, and What.CD provides the map to find these files, the website itself contains no actual

media. Related to this, What.CD does not need storage space to contain the approximately 385,000 albums it tracks at present. Instead, all of these files are stored on over 100,000 users' computers and hard drives. The bandwidth to transfer the data is similarly almost entirely shared by users, with none of the bandwidth burden for actual data transfer shouldered by What.CD.

Chapter 3: Site Governance

3.1 User Classes and Staff

At the core of what makes private BitTorrent trackers successful is their frequent adherence to a system of user classes based on ratio status and other behavior at the site. While a great deal of What.CD's success relies on its openness and freedom regarding user's abilities to create and modify content and information, there is a hierarchy of privileges in the form of a user class structure. A primary purpose of this class system is to generate high-quality content and participation from as many users as possible. Classes create a meritocracy which encourages users abide by site rules and ultimately rewards those users who invest in improving the site through their contributions (O'Mahony 1082). Although the participation and contribution of users is one main source of the tracker's success, to fully comprehend and appreciate the What.CD model it is necessary to understand the role that governance plays at the site.

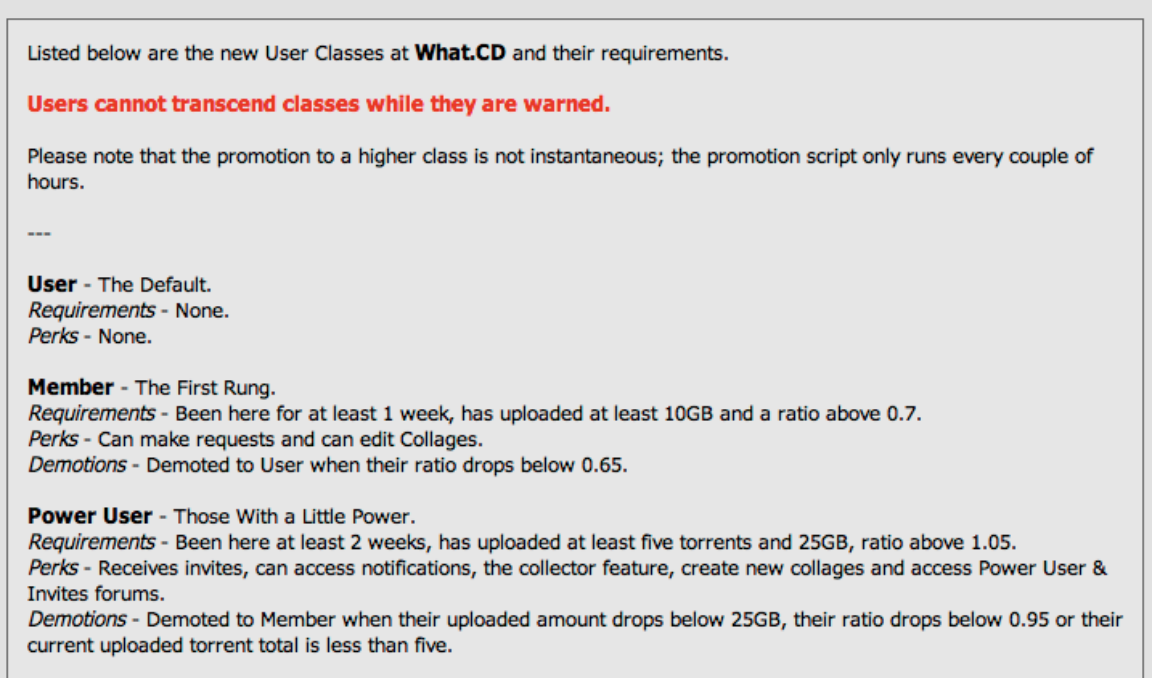
At the top level of the class hierarchy are the site's administrators. To manage the ever-increasing numbers of both torrents and users, an authority is necessary to help keep the site running smoothly and to ensure that the system that has been created continues to function as intended. (O'Mahony 1081). A group of approximately 30 staff members inhabit roles as developers, system administrators, moderators and general support, making them the governors at What.CD. Several of the staff in this category were the original founders of What.CD, and continue to support and help build the site as it grows.

Implementation of new ideas that are frequently requested or suggested by users in the forum is one of the responsibilities these administrators are tasked with.

The lack of a user structure with any type of regulation leads to common issues at many public sites, such as incorrect tagging and metadata information, unreliable or malicious uploads, and a host of other problems that work against content cohesiveness and quality.

Significantly tied to this is the role that trust and reputation play at private, invite-only trackers like What.CD. This hierarchy is helpful to all members of the site as they can look to class indicators for confirmation on the reputation of a given upload. Coupled with the ability to see when a user has joined the site, this distinguishes uploaded content that is of the quality expected at What.CD (Satsiou and Tassiulas 466). It also allows for users to have their reputations denoted in the form of their user class and markers tied to their usernames, such as warnings. Because the nature of a tracker like What.CD is so closely tied to a mindset of collaboration and sharing, it is vital that users feel there is accountability for negative actions and recognition for positives ones, both of which can be reflected in the system of user classes.

User classes signify the level and quality of participation in the community, and What.CD grants or revokes additional privileges as a user's class changes based on his level of participation. Participation is largely measured in ratio, which is discussed in greater detail in Chapter 4. To stay in possession of all site functionality, including the ability to upload new content, users have to reach certain achievements, which include the length of their membership at the site, a minimum ratio level, and for some of the higher classes a minimum number of new uploads.



Listed below are the new User Classes at **What.CD** and their requirements.

Users cannot transcend classes while they are warned.

Please note that the promotion to a higher class is not instantaneous; the promotion script only runs every couple of hours.

User - The Default.
Requirements - None.
Perks - None.

Member - The First Rung.
Requirements - Been here for at least 1 week, has uploaded at least 10GB and a ratio above 0.7.
Perks - Can make requests and can edit Collages.
Demotions - Demoted to User when their ratio drops below 0.65.

Power User - Those With a Little Power.
Requirements - Been here at least 2 weeks, has uploaded at least five torrents and 25GB, ratio above 1.05.
Perks - Receives invites, can access notifications, the collector feature, create new collages and access Power User & Invites forums.
Demotions - Demoted to Member when their uploaded amount drops below 25GB, their ratio drops below 0.95 or their current uploaded torrent total is less than five.

Figure 1. Sample of user classes and specifications; *What.CD;* what.cd/ wiki.php?action=article&id=21*;* 27 Oct. 2007; Web; June 2010.

When a user uploads material or content that is unacceptable at What.CD, or does not follow the rules and guidelines related to acceptable quality standards, they might receive a warning or demotion in user class. Along with the loss of previously mentioned site privileges, the user is expected to demonstrate through positive contribution and participation to have a warning removed and their ability to move up in class reinstated. Failure to get rid of a warning within a designated amount of time may lead to a user being banned from the site entirely.

3.2 Invitations and Access

What.CD is a private, invite-only BitTorrent tracker. More importantly, a new user can only register after being invited to the site by a current member of What.CD. Receiving an invitation to sign-up and register with a username and password is the only way in to What.CD. While there are certainly numerous ways one might go about

seeking out an invite, what is most important to this particular study are the implications of the invite system. Invitation-only systems are both a means to restrict and limit access to the community and files being shared, and as a way of ensuring that those who do gain access are going to more likely be upstanding, contributive members of the site.
The most common method of becoming part of the What.CD community is to be invited by a friend. Knowing another user (whether off or online) is often the simplest way of acquiring an invitation to What.CD.

A second option, for invite seekers without a friend already on the site (or without a friend who has invites to distribute) would be to find their way to the What.CD IRC chat room. This would demand somewhat minimal knowledge of the IRC chat system, and then the ability to prove to site administrators that they are aware of, and familiar with, standard BitTorrent procedures and site guidelines. Another way into What.CD is to ask for and receive an invitation from a user on another invitation-only torrent site. Private torrent sites generally encourage reliable, involved members of their own sites to invite similarly minded and capable users from other sharing communities.

The ability to invite new users to the site is considered a privilege, and as such, comes with its own set of rules and recommendations. Acquiring invitations, which can be distributed via email to prospective users, depends on meeting certain ratio requirements and staying in good standing on the site for a specific duration of time. At milestones along these requirement lines, users will be granted 1 or more invitations, to be used as they see fit. In order to maintain a fairly consistent number of registered users, and to stay at or below the maximum user capacity, invitations are periodically closed. Users who have not logged in to the site for a certain amount of time, or who have

violated rules and lost their accounts will then be pruned, and spots for new members to be invited will be made.

What.CD's private, invite-only status sets it apart as a successful tracker especially when viewed alongside the plethora of public trackers. A major part of what makes What.CD able to succeed in so many areas that other, mostly public trackers do not is that it is private, and open only to those who have been invited to use the site. Looking at some of the major public torrent trackers is necessary to fully understand the uniqueness of What.CD. The scope and organization of What.CD is impressive on its own, but compared to even the best run public trackers, it truly has no equal.

One theory is that behind the gates of a private site, users feel more comfortable sharing larger amounts of copyrighted material. This possibility easily comes under scrutiny, however, when considering the large public trackers that offer hundreds or thousands of movies, television shows, music and software. I argue that it is actually the same accountability that keeps users away from public sites (in terms of what they are uploading, and who might access it) that proves appealing at a private tracker (being held to a standard that is linked directly to their own use of the site). Issues like regulating torrent accuracy, tag accuracy, file quality, etc. do not work nearly as well on sites with open registration, use and visibility, due to the lack of accountability and much higher user numbers (among other issues).

The awarding of invites to current users who can then distribute these invites to potential users is a heavily regulated feature. To receive invitations, a user must first demonstrate that they themselves understand the rules of the site, and prove that they can maintain a status on the site as a contributor and participant. By reaching a certain ratio

(generally somewhere just over 1.0), and having been a member for a designated amount of time (several weeks), one or two invites will then be granted. To gain additional invites, other requirements are set up so that they are given out as long as the user stays in good standing (above a certain ratio) and when certain "data uploaded" waypoints are reached. This provides incentive to users to continue using the site and to maintain good standing in order to become the recipients of additional invitations. Paul Jones explains how a reward (such as access to additional invites, in this case) strengthens a contribution-based online community:

> By giving contributors and readers access to tools for evaluating, ranking, and managing the collections, we are not just off-loading work; we are building communities and intellectual discourse. Strong community members are recognized by reputation capital and trust metrics and are rewarded. ("Open Sourcing" 46)

It should be noted that there is usually a strict policy against the selling of invites, though often invitations are traded for access to other private torrent sites.

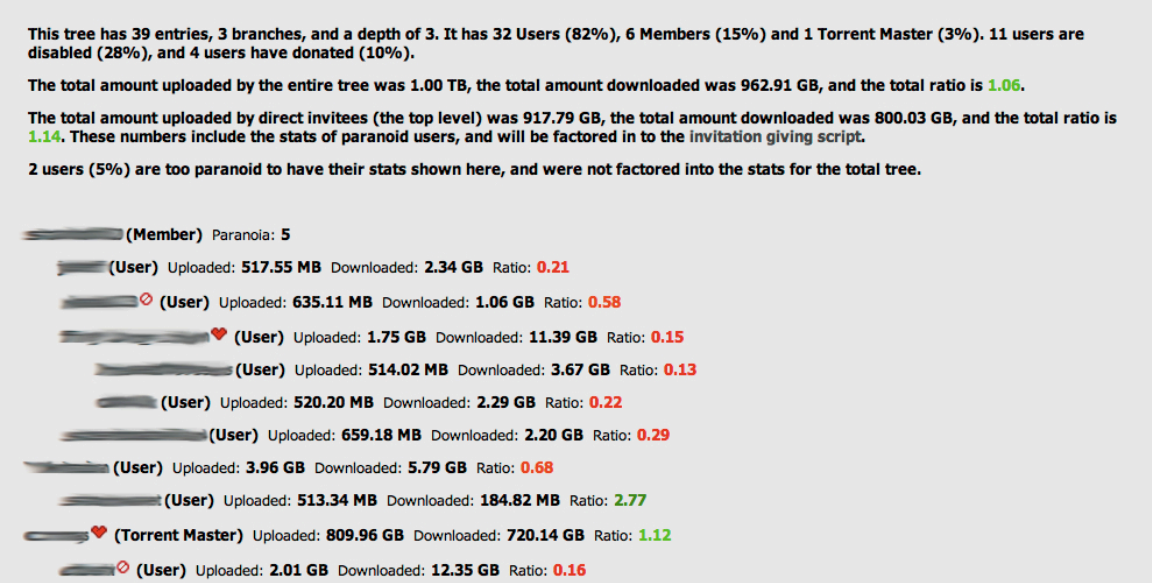


Figure 2. Example of an invite tree; Invite Page; *What.CD;* what.cd/ user.php?action=invitetree; 27 Oct. 2007; Web; June 2010.

While there is certainly an appeal in receiving invites to bring friends to the community or trade for access to others, What.CD offer other incentives to promote careful access. Invite trees chart the history of users and their invitees. If one of the users drops below a certain ratio threshold and eventually ends up losing their account, or if they break the rules in another way to warrant being banned, the top-level user may also be penalized. Conversely, top-level users on the tree may be rewarded with additional ratio or other perks for inviting good members. It is encouraged that prospective invitees be taught the rules of the site, as well as the methodology and philosophy of BitTorrents before they have access to the material.

All of these methods to obtain an invitation work towards creating a trustworthy community of users who will share and follow the rules of the site. Because so much of the success of the site is predicated on honesty and good behavior, restricting access to a public user-base is a necessity. The invite and registration systems work to make each individual user accountable.

Chapter 4: Ratio and Economy

The centerpiece of the sharing model used at What.CD consists of users' upload:download ratio. Nazareno Andrade et al. gives a technical definition of this sharing model as: "The sharing ratio of a peer is the total amount downloaded. By analogy, we define the sharing ratio of an entire torrent at a specific moment to be the total amount uploaded divided by the total amount downloaded by the peers active in the torrent at that moment" ("Influences on Cooperation" 2). At the most basic level, the idea of the torrent system is that users share as much as they take, if not in raw content (new uploads to the site), then in the bandwidth used to transfer files between users (or "peers," in torrent terminology). It makes no difference what a user downloads, as the only measure is the total data transferred downstream. For example, downloading five albums equaling 400 megabytes is no different than downloading one album that is 400 megabytes. As a user takes (downloads) from the site and shares (uploads), a ratio of the uploaded amount to the downloaded amount is calculated. "Since a peer can efficiently increase its sharing ratio by uploading content as a seeder, the mechanism provides an indirect incentive for seeding as well as a direct incentive not to freeride" (Andrade et al., "Influences on Cooperation" 2). This ratio is calculated and updated instantly by the site. Depending on the individual's user class, they will have a required ratio figure to stay in good standing on the site. In this way, uploaded and downloaded data becomes the site's economic unit.

Figure 3. A user's upload: download ratio display; *What.CD;* what.cd/torrents.php*;* 27 Oct. 2007; Web; June 2010.

Nearly all BitTorrent trackers that require a registration – public or private – rely on a ratio system to ensure that material is kept available and accessible. By requiring a minimum amount of upload bandwidth to that which is taken by downloading, trackers can maintain (depending on their ratio requirements) a fairly evenly distributed seeder-base. What this means is that ideally, most torrents will remain seeded by at least one user, so that there are not situations where torrents are inaccessible with no one to leech from. This benefits those users who choose to stay as seeders on low-traffic torrents, as they will see fewer leechers , but receive a higher volume of the uploaded bandwidth. Satsiou and Tassiulas write, "…current incentive mechanisms for BitTorrent-like networks do not motivate peers who already downloaded their files (seeds) to stay in the system and keep on contributing resources to other peers" (468). By relying most heavily on the ratio system as a means to keep seeds alive and connected, What.CD does not have too much of a problem with delinquent users and unusable or dead torrents.

Ratio also solves the free-rider dilemma present in other sharing communities. Because users are kicked off for poor ratio status, What.CD quickly jettisons those users who do not keep to the community rules of contribution and reciprocity. So, What.CD takes on this issue not only through the positive reinforcement of making it beneficial for users not to free-ride, but also creates negative consequences for those who do try to take advantage of the system. "If users who do *not* contribute to a public good – 'free riders' – can benefit from that good on equal terms with those who do contribute, how can one

motivate users to contribute rather than free ride?" (Von Hippel 215). What.CD has answered this question in the form of a variety of incentives and punishments – positive and negative reinforcement – to motivate users to contribute. It has also devised an absolute minimum amount of participation required before a user is deemed a free rider, and is no longer welcome in the community. What this means is that those users attempting to free ride on What.CD, as on most private torrent trackers with ratio requirements, will lose access to the site entirely.

The two key ways to gain upload capital are by contributing new torrents to the site, or by downloading and then seeding existing torrents. The main difference between these two methods is that by adding a new torrent, the user does not expend any download capital; he only gains whatever is downloaded from his torrent. Downloading an existing torrent will lower the user's ratio in the amount of the torrent's file size, but anything uploaded from that torrent (whether during the download, or after during the seeding process) goes back towards the upload capital (a seed is a completed torrent made available by at least one user). Over time, the more completed torrents (seeds) a user makes available, and the more times those seeds are utilized by other users, the higher the overall upload amount. There are other ways, however, to affect one's ratio. As will be discussed in Chapter 5, filling requests and collecting bounty is another way of adding upload capital instantly to one's ratio.

The ability to see which users snatched any torrent at What.CD is helpful, so that should a user find an unseeded torrent, he can private message several members on the "snatched" list, and request that someone begin seeding the torrent again. Once the seed is started, the lecher can begin to download the torrent. Being granted bonuses for

maintaining good upload/download ratio is a key feature of What.CD, and serves as one of many incentives to participate by seeding as many torrents as possible.

What.CD takes many efforts to help the community stay cohesive and active by instituting various ways to encourage and reward users with several reward type features. Furthermore, these same systems help create a healthy and robust system of sharing to keep the core mechanisms of the site (like ratio) from sinking too low to repair. This section will explore a few of the implementations What.CD has undertaken.

When "cooperation levels" are lagging, administrators have the ability to stimulate cooperation (Andrade et al., "Influences on Cooperation" 1). A "freeleech period" is occasionally implemented by administrators at What.CD, where either all, or just select torrents, are made free to download at no cost to a user's upload ratio. Their upload contributions are counted and added, however, giving a boost to all users' ratio. This serves to both reward users and to increase the seeders on many older, under-seeded torrents. At other times, freeleech may enacted be to celebrate a holiday (freeleech has been turned on around every Christmas and New Years since What.CD was started). In this instance, it is clearly to reward users and is seen as a gift. It also allows users who may have fallen behind on their ratio upkeep to lift their status significantly more easily.

Another reason freeleech may be used is in the promotion of certain artists or albums by staff members. Featured music may be given freeleech status to help drive interest towards new or unknown artists, sometimes even those who are What.CD users themselves. Other times, staff selections will incentivize users to try new genres or artists, at no cost to their ratio. Witten et al. write:

> A library, by its very nature, is both developed by a community and exists to support the activities of a community of users. Digital libraries have

> great potential to enable communities to further influence and extend their audiences. (124)

The What.CD torrent tracker works in a similar fashion, bringing exposure to the varied items in a vast music collection and the influence of users' recommendations to one another. This is another of the unique and forward-looking features of What.CD. As a way to both distribute music and keep community participation up, the promotion of unknown artists who are What.CD members serves a dual purpose.

Chapter 5: Site Features

Previous chapters detailed What.CD user features such as invitations and classes. This section will focus on more of the content-focused features found on the tracker. Perhaps the most fundamental tenet of What.CD is that the material and the way it is presented on the site should be as accurate and of the highest possible quality. File format and bit rate standards are firmly outlined, indicating what the format of files and the minimum quality allowance is for each type. Coupled with this are detailed instructions and guides to using specific compact disc ripping applications, configured with specific settings, for accurate and error-free digital files. It is encouraged that proof, in the form of .log and .cue files, be provided along with the music files when uploaded to the tracker. These restrictions and guidelines are one of the huge differences between a private site like What.CD and public torrent sites, and due to the excellent quality standards enforced at What.CD, it achieves greater success in both material available and user participation. “The pressure of being responsible to one’s peers and to be corrected by them, if in error, might be one key factor for the high degree of reliable metadata. However, a pride in one’s collection and a willingness to share also contribute to the quality and reliability of the data and the metadata” (Jones, “Strategies and Technologies” 659). At the upload stage, a vital step in maintaining the superiority of a private community is found. The fields that require information to be provided are

extensive and specific. Starting with the basics of the Artist Name and Album Name, there is then a plethora of other data tags to assign.

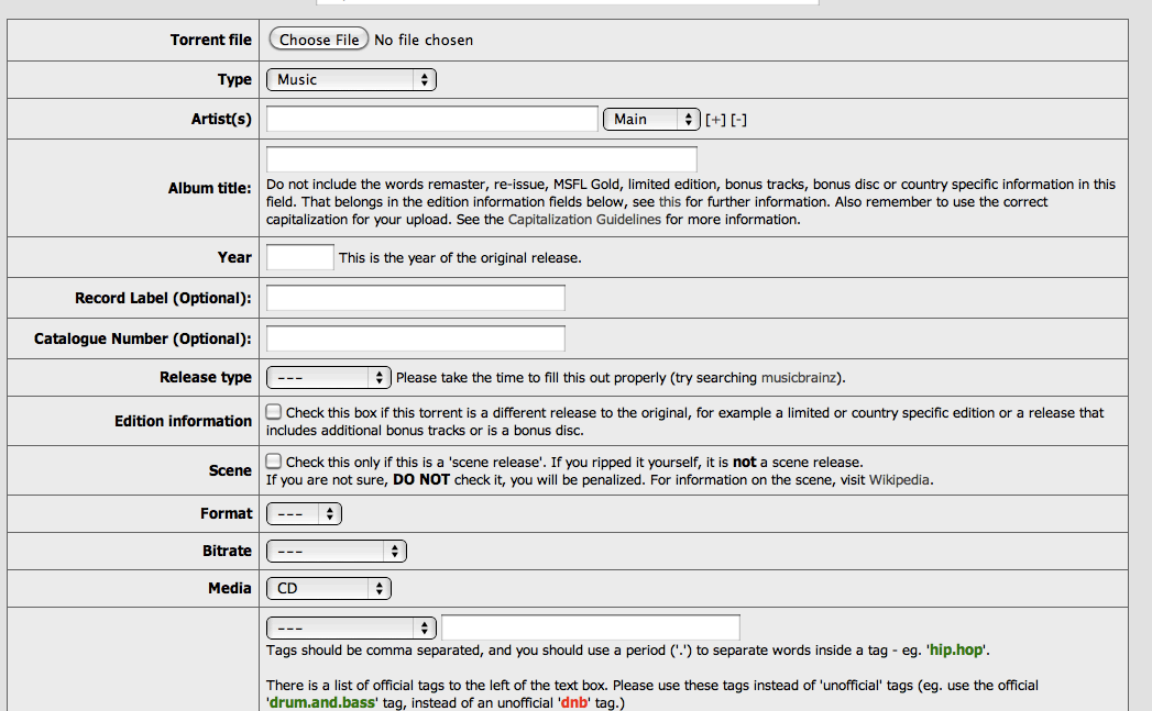
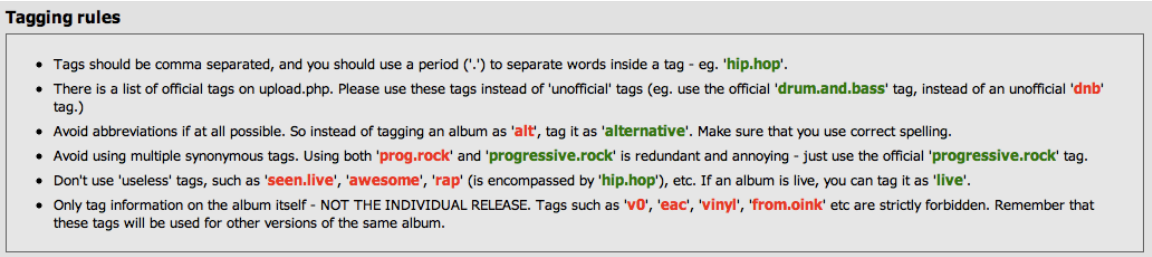


Figure 4. Upload page at What.CD; *What.CD;* what.cd/upload.php; 27 Oct. 2007; Web; June 2010.

The benefits to having these requirements are multifold. At the base level, they serve to keep the original uploader somewhat honest at the initial stage of providing material to the site. By having to go through several steps of defining what it is they are uploading, it becomes harder to simply upload anything. Mistakes in these fields will be identified quickly as being incorrect by downloaders or administrators. Secondly, the information provided by these fields once the material is uploaded allows potential downloaders to get the best possible idea of what exactly the specific torrent will be providing. If for instance, a user already has an album in physical CD format, they may also want a vinyl record rip of that same album, for a slight difference in sound. Alternatively, perhaps they had originally ripped their own CD at a low quality bit rate,

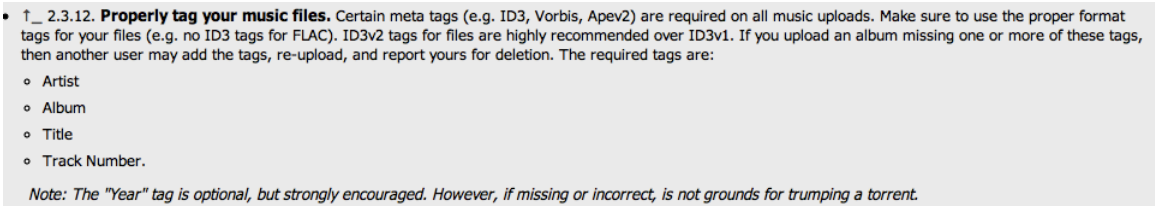
and decide (whether because of a lost CD or inability to produce a quality rip themselves) they want a version at a higher bit rate. These tags then act as the sole way of a user actually knowing what the file he intends to download will be.



Tagging rules

- Tags should be comma separated, and you should use a period ('.') to separate words inside a tag - eg. '**hip.hop**'.
- There is a list of official tags on upload.php. Please use these tags instead of 'unofficial' tags (eg. use the official '**drum.and.bass**' tag, instead of an unofficial '**dnb**' tag.)
- Avoid abbreviations if at all possible. So instead of tagging an album as '**alt**', tag it as '**alternative**'. Make sure that you use correct spelling.
- Avoid using multiple synonymous tags. Using both '**prog.rock**' and '**progressive.rock**' is redundant and annoying - just use the official '**progressive.rock**' tag.
- Don't use 'useless' tags, such as '**seen.live**', '**awesome**', '**rap**' (is encompassed by '**hip.hop**'), etc. If an album is live, you can tag it as '**live**'.
- Only tag information on the album itself - NOT THE INDIVIDUAL RELEASE. Tags such as '**v0**', '**eac**', '**vinyl**', '**from.oink**' etc are strictly forbidden. Remember that these tags will be used for other versions of the same album.

Figure 5. Basic tagging rules of the site; *What.CD;* what.cd/rules.php?p=tag; 27 Oct. 2007; Web; June 2010.



- ↑_ 2.3.12. **Properly tag your music files.** Certain meta tags (e.g. ID3, Vorbis, Apev2) are required on all music uploads. Make sure to use the proper format tags for your files (e.g. no ID3 tags for FLAC). ID3v2 tags for files are highly recommended over ID3v1. If you upload an album missing one or more of these tags, then another user may add the tags, re-upload, and report yours for deletion. The required tags are:
  - Artist
  - Album
  - Title
  - Track Number.

Note: The "Year" tag is optional, but strongly encouraged. However, if missing or incorrect, is not grounds for trumping a torrent.

Figure 6. Upload tagging rules; *What.CD;* what.cd/rules.php?p=upload; 27 Oct. 2007; Web; June 2010.

This detailed tagging and labeling system serves other purposes, as well. Another of the many benefits to What.CD is that duplicate material is non-existent. For example, users do not have to decide between two of the same albums of the same format or at the same bit rate. By employing detailed tags and labels, it is simple to verify what is already available thereby drastically reducing duplicate uploads.

Finally, these tagging mechanisms work for the very reason that they were originally designed: advanced searching and filtering. As opposed to the very limited and often useless search features of many public sites, at What.CD, search parameters are incredibly detailed and flexible. In addition to searching by any of the fields indicated in the section above, users have the option of searching based on torrent activity (number of

peers, seeders, leechers, comments, etc.), number of files in a torrent, data size of a torrent, the user who uploaded the torrent, when the torrent was added, and several other specific options.

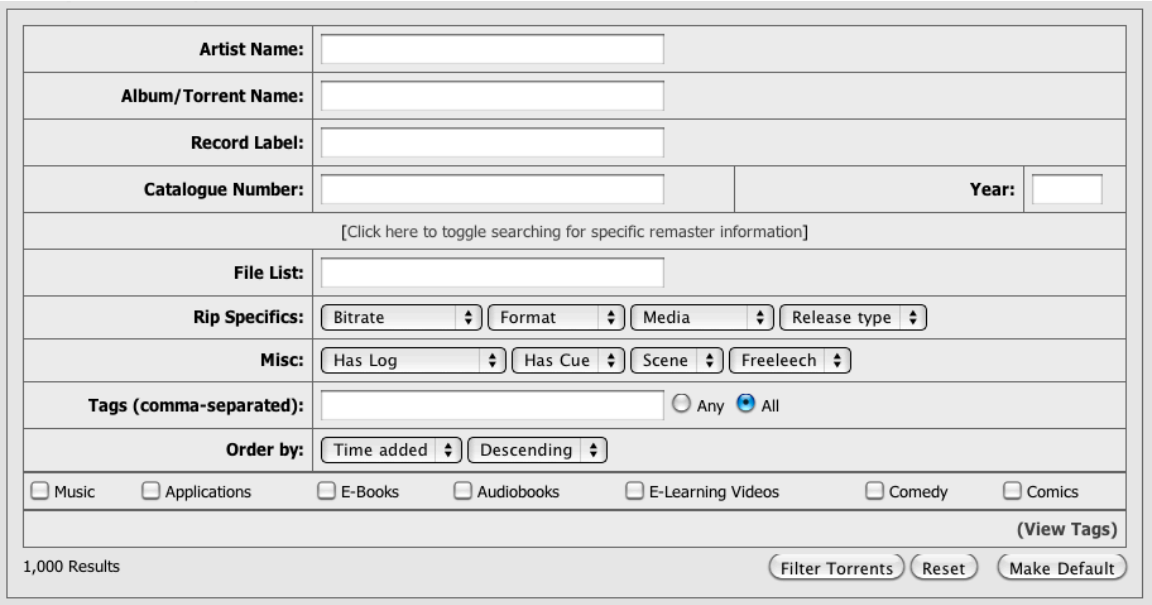


Figure 7. Advanced search filters; *What.CD;* what.cd/torrents.php?action=advanced; 27 Oct. 2007; Web; June 2010.

In addition to the search functionality, tag clouds and "related artists" features are prominently displayed along with each album, and like the search, are only possible because of the detailed tagging required for every upload. All of these components contribute to the advanced capabilities found only on What.CD or other strictly maintained private torrent communities. Attempts to include some of these features can be seen at some public sites, but rarely does the end result come close to the quality and usability of a private tracker.

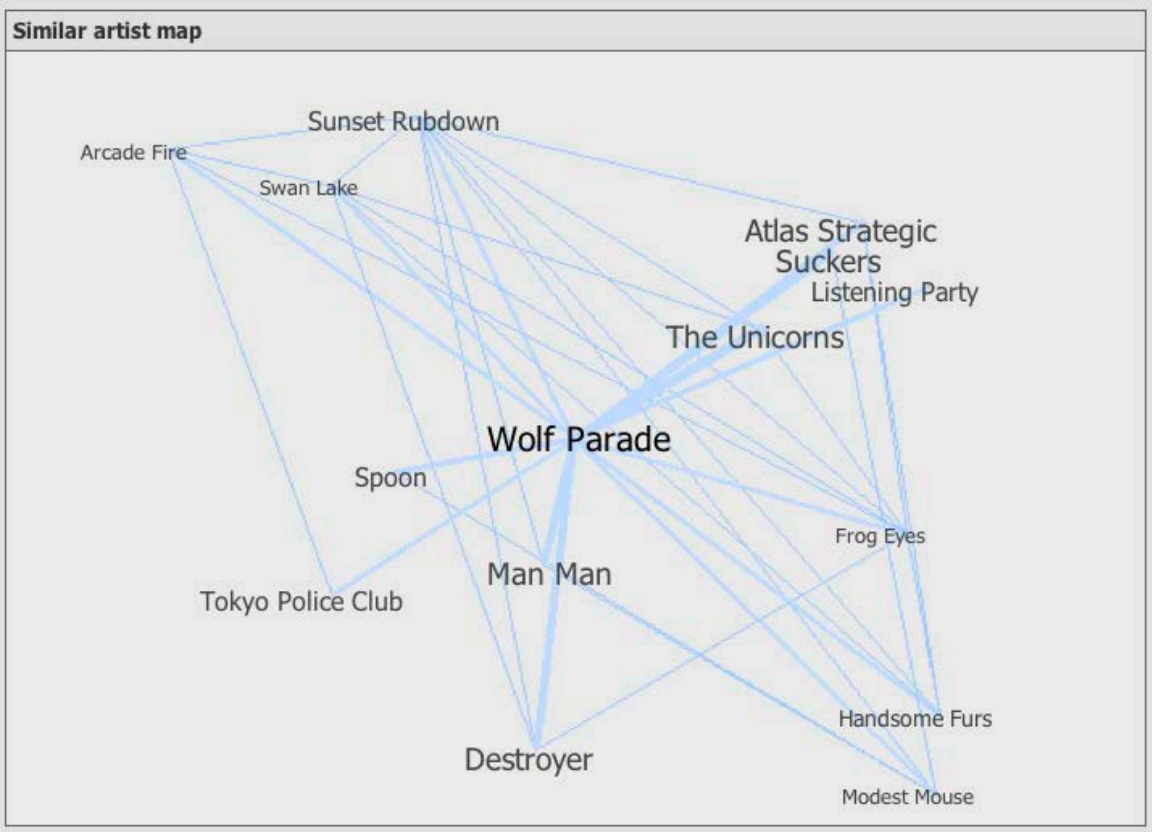
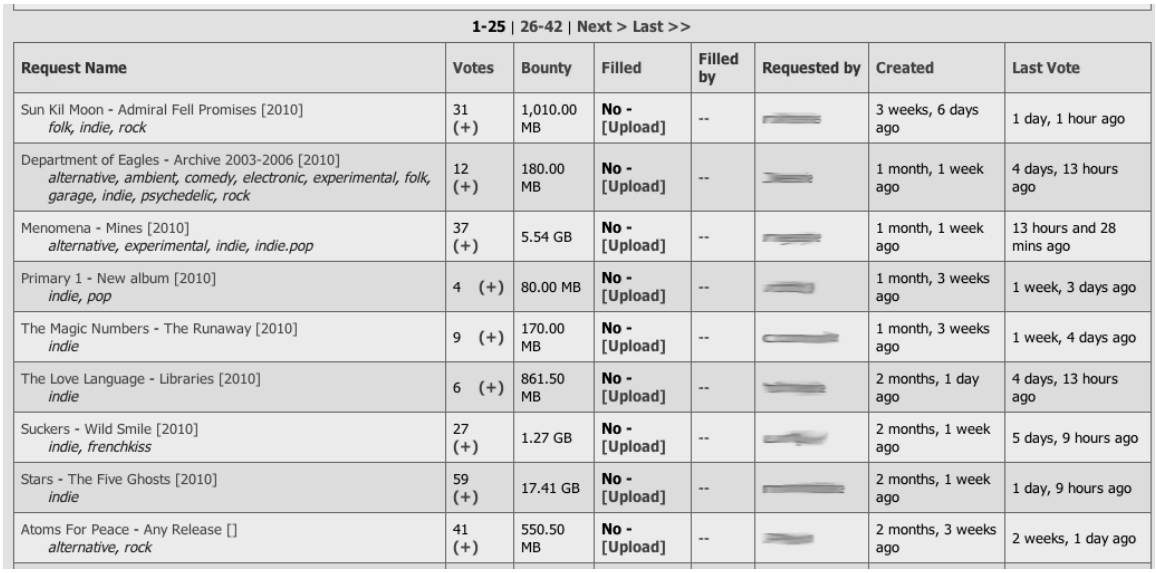


Figure 8. Similar artist map; *What.CD;* what.cd/artist.php?id=1069*;* 27 Oct. 2007; Web; June 2010.

Yet another unique feature found on What.CD is the request system. Users have the ability to make specific requests for material that is not yet available on the tracker. In many cases, these requests are for rare albums that are hard to find for purchase, advanced unreleased albums, and in some cases, overlooked albums that are simply just not yet on the site. The request system is a key draw to using a site like What.CD, as it allows for the user to literally ask for and – in time – receive. Making a request is as simple as entering the artist name and album title. The next step, however, allows for other users to weigh in on the demand and value of a specific request.

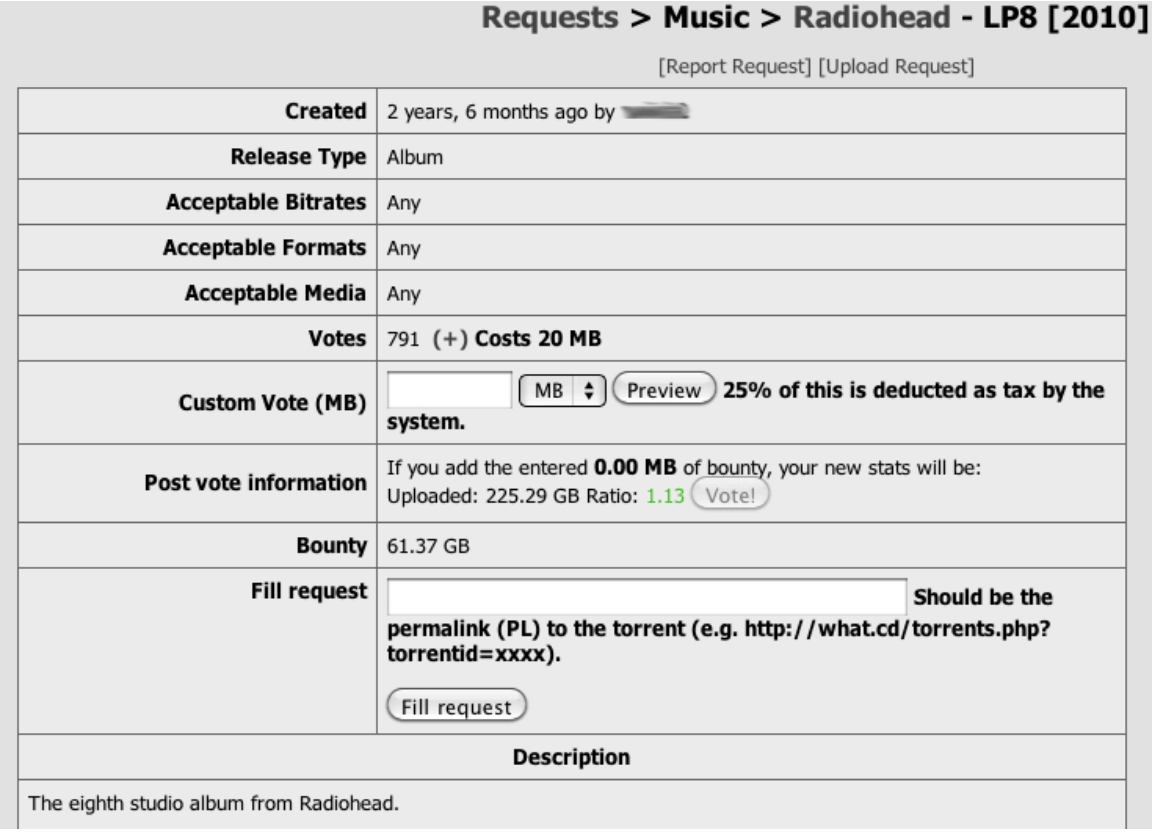


1-25 | 26-42 | Next > Last >>

Request Name	Votes	Bounty	Filled	Filled by	Requested by	Created	Last Vote
Sun Kil Moon - Admiral Fell Promises [2010] *folk, indie, rock*	31 (+)	1,010.00 MB	**No - [Upload]**	--	[redacted]	3 weeks, 6 days ago	1 day, 1 hour ago
Department of Eagles - Archive 2003-2006 [2010] *alternative, ambient, comedy, electronic, experimental, folk, garage, indie, psychedelic, rock*	12 (+)	180.00 MB	**No - [Upload]**	--	[redacted]	1 month, 1 week ago	4 days, 13 hours ago
Menomena - Mines [2010] *alternative, experimental, indie, indie.pop*	37 (+)	5.54 GB	**No - [Upload]**	--	[redacted]	1 month, 1 week ago	13 hours and 28 mins ago
Primary 1 - New album [2010] *indie, pop*	4 (+)	80.00 MB	**No - [Upload]**	--	[redacted]	1 month, 3 weeks ago	1 week, 3 days ago
The Magic Numbers - The Runaway [2010] *indie*	9 (+)	170.00 MB	**No - [Upload]**	--	[redacted]	1 month, 3 weeks ago	1 week, 4 days ago
The Love Language - Libraries [2010] *indie*	6 (+)	861.50 MB	**No - [Upload]**	--	[redacted]	2 months, 1 day ago	4 days, 13 hours ago
Suckers - Wild Smile [2010] *indie, frenchkiss*	27 (+)	1.27 GB	**No - [Upload]**	--	[redacted]	2 months, 1 week ago	5 days, 9 hours ago
Stars - The Five Ghosts [2010] *indie*	59 (+)	17.41 GB	**No - [Upload]**	--	[redacted]	2 months, 1 week ago	1 day, 9 hours ago
Atoms For Peace - Any Release [] *alternative, rock*	41 (+)	550.50 MB	**No - [Upload]**	--	[redacted]	2 months, 3 weeks ago	2 weeks, 1 day ago

Figure 9. Request listings; *What.CD;* what.cd/requests.php*;* 27 Oct. 2007; Web; June 2010.

All users have the ability to indicate that they also want a particular item by clicking on a "vote" option, which then tallies the users requesting that material. There is also a second way to voice varying degrees of demand for a particular request. By offering "bounty," in the form of a specified ratio amount, users can contribute their own ratio to whoever fills (uploads) the request. In doing so, requests are distinguished not only by how much of the populous wants the item but also, by how much desire there is for an item within the group of users. This also helps ensure that the request system is not abused, as requests are not free. The recurring theme here is you get what you pay for (with "pay" being time, effort and ratio).



Requests > Music > Radiohead - LP8 [2010]

[Report Request] [Upload Request]

Created	2 years, 6 months ago by [redacted]
Release Type	Album
Acceptable Bitrates	Any
Acceptable Formats	Any
Acceptable Media	Any
Votes	791 (+) **Costs 20 MB**
Custom Vote (MB)	MB Preview **25% of this is deducted as tax by the system.**
Post vote information	If you add the entered **0.00 MB** of bounty, your new stats will be: Uploaded: 225.29 GB Ratio: 1.13 Vote!
Bounty	61.37 GB
Fill request	**Should be the permalink (PL) to the torrent (e.g. http://what.cd/torrents.php?torrentid=xxxx).** Fill request

Description

The eighth studio album from Radiohead.

Figure 10. An album request, vote and bounty page; *What.CD;* what.cd/requests.php?action=view&id=5831; 27 Oct. 2007; Web; June 2010.

The requests system also benefits the user who has the ability to fill a request. The material already present may overwhelm a new member to the site, and the request system makes it simple to check what is not currently available. Furthermore, a user hesitant to upload material may be coerced by the visible reward of upload ratio to be gained by filling a particular request. As opposed to a public site that offers no tangible reward for providing new material, What.CD privileges the uploader with new resources to download material for themselves.

This chapter has described only a few of the major features utilized at What.CD. The tracker is constantly evolving, adding and refining functionality to enhance the user-experience and improve the library of material. All of the features ultimately serve a

function to establish a meritocracy in the community and to keep up overall quality of content at the site.

Chapter 6: Community, Collaboration and Contribution

Regardless of users' motives in using the site, it is clear that it has stretched beyond the typical idea of a website used to steal music. The intent of most users is not merely to find the single they heard on the radio, or to literally rob others of their work. Instead, it is out of a passion for collecting and listening to music, and What.CD is a simple one-stop source, with near on-demand access to more music than any single store or library could be expected to house. Chun-Yao Huang writes:

> …the popularity of music file sharing among certain consumers around the world can be taken as the emergence of a new subculture or of a new mode of consumption that represents 'tensions between the market and (sub-cultural) consumption communities' that cannot be explained by traditional utilitarian motivations. (40)

Furthermore, even online digital music stores like Itunes or eMusic either don't have the rights, or do not carry all of the material than can be found on What.CD. Familiar filters like popularity, release date, related artists, etc. are all retained in the What.CD system, and are arguably more accurate due to the concentrated tastes of What.CD users.

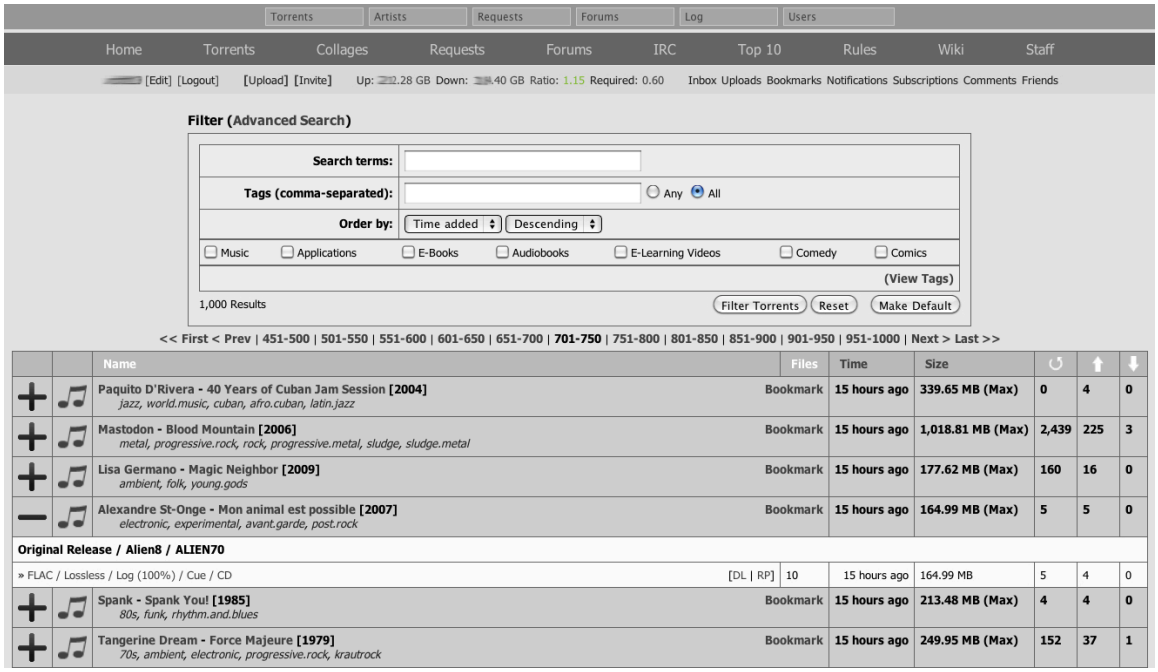


Figure 1. Primary torrent browsing page; *What.CD;* what.cd/torrents.php; 27 Oct. 2007; Web; June 2010.

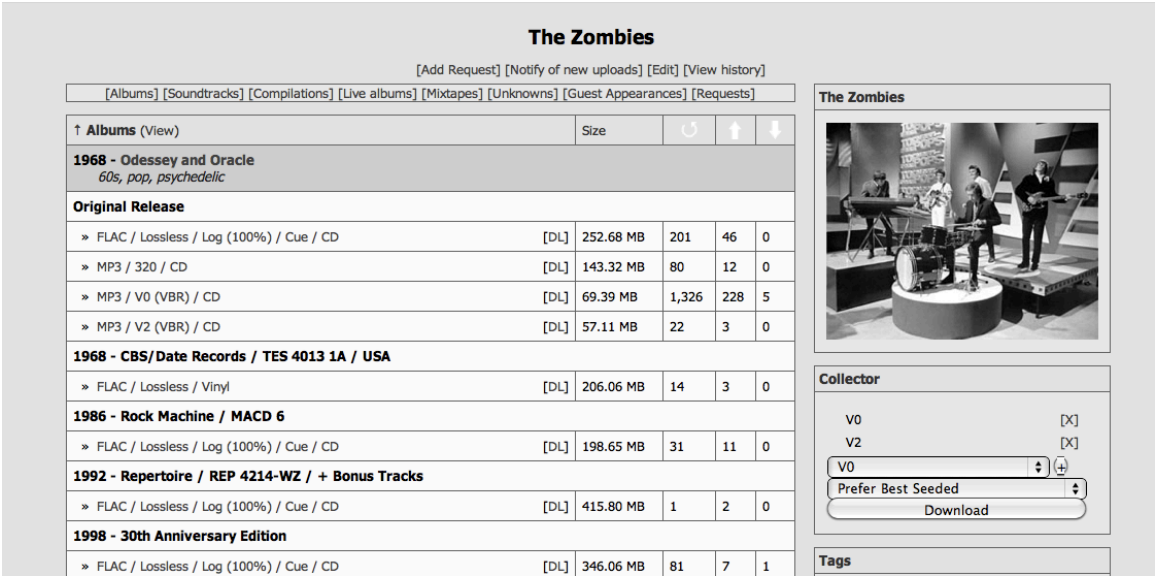


Figure 2. An artist's page at What.CD; *What.CD;* what.cd/artist.php?id=15636; 27 Oct. 2007; Web; June 2010.

Contrasting most paid services, where music is offered by companies who have deals with record labels, or buying direct from record labels themselves, What.CD allows

users to see what type of music is being uploaded by whom. This allows for greater oversight of the clustering of musical taste, and the habits of listeners. In addition, because of the ease of acquiring music through What.CD, users are afforded the ability to branch-out and experiment with new listening choices that they might not otherwise experience (due to lack of money, lack of access in the form of record stores, etc.). It is one of the most attractive elements of a site like What.CD.

Related to the community framework aspect is the fact that members of What.CD seem to overwhelmingly work together towards a common goal – that of building a massive, well-organized music archive. There are many examples of web experiments, where sites evolved or devolved based upon the behavior of members (Myspace, eBay, etc.). What.CD happens to be an example of a site guarded behind a wall of invite-only access that has for over two years stayed positive and cohesive. This invites another question, of whether or not (or just how much) the exclusive nature of What.CD helps the spirit of "teamwork."

What emerges from all of this is that users become more able to navigate a website that offers more complex search features and navigational tools. It contains a much more vast collection of music, in far more shapes and sizes than other commercial systems. It becomes a way not to steal music, but to have a search engine that leads to the music from all times and genres. Huang writes:

> Moreover, in a study of online communities, Evans, Wedande, Ralston, and Hul concluded that experienced users are more likely to participate in a virtual community site primarily as a source of information and to discuss topics of interest, whereas novice users are particularly attracted to the interaction it offers with other people and the opportunity to make new friends. To explain this difference, the authors argued that there might be a shift from social usage to a more functional usage of virtual communities as one becomes more experienced. (41)

Clearly the appeal of What.CD as only a destination to steal music evaporates, if not at the very start then soon after. The collection and catalog then take over, along with the social framework the site employs. As previously mentioned, the digital library becomes an environment where large volumes of data and information can be browsed and utilized by huge numbers of users, or digital "visitors."

The idea of community and participation as a keystone of the success of What.CD has been a thread throughout the paper. While What.CD still maintains a plethora of rules, guidelines, and restrictions, the site is still unburdened by some of the typical constraints of media distribution systems (legality, as one strong example). Acting as a repository for content distributed globally, What.CD brings this content into a well-organized, searchable archive of digital music. As a result of its large user base, located in dozens of countries, with access to hundreds of thousands of music releases, it has become the largest music archive found anywhere. An argument could be made that the motivating factor behind the majority of users' use of the site is not to get music for free. Instead, it is out of a passion for collecting and listening to music, and What.CD is a simple one-stop source, with near on-demand access to more music than any single store or library could be expected to house. Furthermore, even online digital music stores like Itunes or eMusic either don't have the rights, or do not carry all of the material than can be found on What.CD. Familiar filters like popularity, release date, related artists, etc. are all retained in the What.CD system, and are arguably more accurate due to the more concentrated and perhaps discerning tastes of What.CD users.

In the paper, “An Ethnographic Study of Music Information Seeking: Implications for the Design of a Music Digital Library,” Cunningham et al. make the following assertion:

> It will be difficult for a virtual digital library to match the rich collaborative browsing environment afforded by the physical CD store. In stores, friends can impulsively decide to indulge in music shopping, quickly share finds with each other both orally and visually (by holding up a CD or passing it from one person to another), and just as casually end the excursion. Synchronizing collaborative sessions online may be a more bothersome task than simply zipping into a store as friends stroll down a street, although facilities such as instant messaging or chatrooms may provide promising ways to add spontaneity to collaborative exploration of digital libraries. (6)

As recently as 2003 it was being argued that it would be hard to replicate the atmosphere found in physical spaces in a digital medium. The emergence of private torrent sites like What.CD may challenge this belief.

Chapter 7: Conclusion

Previous chapters outlining the features and governance of What.CD have not been given solely to provide dry recitations of information. Instead, the author hopes that a clear and deliberate pattern has become noticeable; that What.CD is a vibrant digital library – a community formed through the accumulation of the previously described disparate parts. It is important to keep in mind that this paper has focused on one perspective of a private torrent tracker, but hopefully raises many other questions to be further explored in future studies. The potential for commercial and business models using the What.CD framework would be an especially complex and fascinating endeavor. Another study might focus on ways in which musicians (independent and on major labels) could leverage a large community of music consumers like What.CD to promote their work. It might also concentrate on those musicians who have already been using the torrent distribution model, and analyze the successes or failures it has brought to those who have tried it.

It is the author's hope that this study was able to illuminate the workings of an environment that demonstrates the viability and strength of a user-created digital library. As national and international law continues to evolve and shape the digital landscape -- especially with the growing controversy over network neutrality -- the way that internet users collaborate and form communities will change, as well. Evaluating other communities and enclaves of digital sharing and collaboration will be vital to better

understanding and creating new, exciting ways of obtaining and disseminating information, data and media in the future.

Bibliography

Andrade, Nazareno, Elizeu Santos-Neto, Francisco Brasileiro, Matei Ripeanu. "Resource Demand and Supply in BitTorrent Content-sharing Communities." *Computer Networks: The International Journal of Computer and Telecommunications Networking* 53.4 (2009): 515-527.

Andrade, Nazareno, Miranda Mowbray, Aliandro Lima, Gustavo Wagner, Matei Ripeanu. "Influences on Cooperation in BitTorrent Communities." *Proceedings of the ACM SIGCOMM Workshop on Economics of Peer-to-peer Systems.* August 25, 2005, Philadelphia, Pennsylvania. 111-115.

Cunningham, Sally Jo, Nina Reeves, and Matthew Britland. "An Ethnographic Study of Music Information Seeking: Implications for the Design of a Music Digital Library." *JCDL '03: Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries.* May 27-31, 2003, Houston, Texas. 5-16.

Huang, Chun-Yao. "File Sharing as a Form of Music Consumption." *International Journal of Electronic Commerce* 9.4 (2005): 37-55.

Jones, Paul. "Open (Source)ing the Doors for Contributor-run Digital Libraries." *Communications of the ACM* 44.5 (2001): 45-46.

Jones, Paul. "Strategies and Technologies of Sharing in Contributor-Run Archives." *Library Trends* 53.4 (2005): 651-662.

O'Mahony, Siobhan, and Fabrizio Ferraro. "The Emergence of Governance in an Open Source Community." *Academy of Management Journal* 50.5 (2007): 1079-1106.

Pomerantz, Jeffrey, and Gary Marchionini. "The Digital Library as Place." *Journal of Documentation* 63.4 (2007): 505-533.

Satsiou, Anna, and Leandros Tassiulas. "Reputation-based Resource Allocation in P2P Systems of Rational Users." *IEEE Transactions on Parallel and Distributed Systems* 21.4 (2010): 466-479.

Von Hippel, Eric, and Georg Von Krogh. "Open Source Software and the 'Private–Collective' Innovation Model: Issues for Organization Science." *Organization Science* 14.2 (2003): 209-223.

What.CD. N.p., 27 Oct. 2007. Web. June 2010.

Witten, Ian, Matt Jones, David Bainbridge, Polly Cantlon, Sally Jo Cunningham. "Digital Libraries for Creative Communities." *Digital Creativity* 15.2 (2004): 110-125.